Erin K. Moore. How does Cognitive Ability impact the use of Query Reformulation Moves? A Master's Paper for the M.S. in IS degree. August 2015. 27 pages. Advisor: Diane Kelly

People have different mental strengths and weakness, which can be measured according to cognitive ability. Learning about strengths and preferences in terms of search behavior, and looking for patterns between behaviors and cognitive abilities, creates the opportunity to make search tools and systems more effectively meet user needs and preferences. While we know that different cognitive abilities exist, and that people form and reform search queries in a variety of ways, we do not know how these two elements interact, or if the interaction is predictable or significant. This paper performs secondary analysis of data collected during a study of cognitive ability, adding in the element of query reformulation moves. It assesses the effect of these cognitive abilities on study participants' search formulation behaviors. Analysis showed that the most common search move was adding a concept to a query, followed by deleting concepts and manipulating search terms. Of the cognitive abilities, the only statistically significant differences between high and low groups were found in the visualization ability. Those in the high skill group made significantly more moves, and significantly more term manipulation moves, than their low skill counterparts.

Headings:

Online Information Retrieval

Query Formulation – Cognitive Aspects

Information Seeking Behavior

Search Strategies

HOW DOES COGNITIVE ABILITY IMPACT THE USE OF QUERY REFORMULATION MOVES?

by
Erin K. Moore

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Information Science.

Chapel Hill, North Carolina

August 2015

Approved by

____

Diane Kelly

Table of Contents

Introduction

This paper presents a secondary analysis of previous work on cognitive abilities and search behaviors, adding in the element of search moves and reformulation patterns. The goal is to identify preliminary relationships between various cognitive abilities and common search reformulation moves.

A number of things impact how users approaches search. One factor explored in this and other studies is cognitive ability level. Cognitive ability is more than just IQ – there are various abilities (e.g., memory, perception, special orientation) each of which can be measured with different tests, and which are useful in different scenarios. A variety of cognitive abilities could impact search behavior differently. By testing for these cognitive abilities before issuing search tasks, researchers make it possible to begin looking for patterns and relationships between cognitive ability and query reformulation. Research in this area could lead to improved search engine interfaces, better understanding of how human differences impact search skill and comfort, and advancement in creating tools and systems that better meet the varied needs of *all* users.

When search tasks are assigned to search study participants, as with the study being analyzed in this paper, those participants go on to formulate queries. While they occasionally submit just a single query, more often they adjust their query based on their satisfaction with the search engine results page (SERP). "Users frequently modify a previous search query in hope of retrieving better results. These modifications are called query reformulations or query refinements” (Huang and Efthimiadis, 2009, p. 77). By studying these reformulations, researchers hope to better understand user search behavior, and to eventually improve search interfaces.

The queries recorded in search studies can be analyzed a number of ways: number of words per query; length of time spent typing, or on the SERP, or browsing results; number of queries per task. Those queries can also be further subdivided, coded by move. A search move is generally a smaller unit than a query. Issuing a query like “how big is America” followed by “united states total square miles” shows several changes occurring at once. Users can get more specific (“big” to “total square miles”), more general, or just try synonyms (e.g. “America” to “United States”) to improve their search results. Studying these behaviors may help to improve search engine performance, and to create systems that better serve the humans using them.

This paper aims to build upon previous work on cognitive abilities and new work on search reformulation moves to discover new relationships to explore in greater depth. The following questions will be addressed:

R1: What is the relationship between cognitive ability and number of query reformulation moves?

R2: To what extent do the frequency distributions of reformulation move (by type) differ between high and low cognitive ability groups?

Literature Review

1.1 Cognitive Abilities

Cognitive abilities are defined as being “comprised of higher mental functions such as reasoning, remembering, understanding, and problem solving” (Carroll, 1993). In the realm of information retrieval and search behavior, these mental functions control the “intellectual processes engaged during search” (Brennan et al., 2014). The strength or weakness of one’s various cognitive abilities may affect not only how effective one’s search is, but *how* one searches. The three cognitive abilities assessed in this paper are associative memory, perceptual speed, and visualization ability.

The Ekstrom Kit of factor-referenced cognitive tests was used to determine the cognitive scores used in this paper. Ekstrom defines associative memory as “the ability to recall one part of a previously learned but otherwise unrelated pair of items when the other part of the pair is presented” (Ekstrom et al., 1976). While studying query formulation, Gwizdka (2009) found that memory span effected performance interacting with word clouds. Twenty-three participants performed four search tasks each, using two different search interfaces. Gwizdka looked at both memory span and verbal closure. He found that those with higher memory and verbal abilities expended more effort in demanding search situations, but that their task outcomes were not significantly better than those of low ability. In a study of both cognitive ability and age, and their effect on information retrieval, Westerman et al. (1995) found that “spatial memory and logical reasoning scores were negatively correlated with information retrieval response times" in a hierarchical database.

Perceptual speed involved "speed in comparing figures or symbols, canning to find igures or symbols, or carrying out other very simple tasks involving visual perception" (Ekstrom et al., 1976, pp.123). In search, this might manifest as skill at browsing results and scanning for relevance. In an early study of an IR system designed to improve subject descriptions in a database of reference abstracts, Allen (1994) found that students with high perceptual speed had superior search performance to those with low perceptual speed. This may have been specific to the system being studied – current search engine interfaces differ significantly from the databases of the early 1990s. Al-Maskari and Sanderson (2011) looked at the effects of search experience and cognitive skill in performing TREC searches. They found that users with lower perceptual speed were less effective in their searches whereas those with higher perceptual speed found relevant items more quickly. They did not look into reformulation of queries, just time to first relevant result.

The Ekstrom manual defines visualization as "the ability to manipulate or transform the image of spatial patterns into other arrangements" (Ekstrom et al., 1976). Campagnoni and Erlich (1989) found that the navigation and editing they had their participants perform were highly correlated with memory and visual abilities. Those with better visualization skills were faster when they retrieved information. They also went back to the top of the information hierarchy less frequently; perhaps maintaining a mental model of the organization structure with more effectiveness than those with low visualization skill. In a study of 35 college students, Downing, Moore, and Brown (2005) had each participant perform five searches using the FirstSearch archival tool. The participants were scored for visualization ability before searching. Downing et al. observed that those with higher visualization ability were faster and found more relevant results than those with lower visualization ability. A larger study (n=101) published a year later by Pak, Rogers, and Fisk (2006) had less definitive results. Visualization did not have consistent effects on search performance in their study, though "spatial orientation ability was related to performance with tasks that were high in their navigational requirement" (p.154).

1.2 Query Reformulation

Despite the evolution of increasingly complex information retrieval systems, user behavior remains an important part of information retrieval. While the results a search engine provides given a specific query are interesting, how the user reacts and adjusts based on those results is an important research area. Query formulation and reformulation have been studied both in the lab and through large-scale naturalistic logs.

Spink, Jansen, and Ozmultu (2000) used a large Excite dataset to look at both query reformulation as well as the impact of relevance feedback on user behavior. At the time, Excite's SERP had a "more like this" button, which Spink et al. used as a measure of relevance feedback. They found that the "more like this" button was not frequently used. Ten years later, Excite is no longer a major search engine, and the "more like this" button is no longer a part of their SERP. They identified term replacement as the most common reformulation move in their dataset.

Six years later, Rieh and Xie (2006) used the same Excite dataset to look more closely at query reformulation. They manually reviewed sessions of six or more queries, with a total sample of 313 search sessions. Unlike other researchers in this area, they looked only at queries, not at sites visited. They had a three-category conceptualization of reformulation: content (changing the meaning of the query), format (adjusting spelling, acronyms, etc.), and resource (e.g., when users shifted between web results, images, and multimedia). They found that 80% of query modifications were content-cased, about 14% were format modifications, and the remaining 5% were resource changes or could not be coded.

Huang and Efthimiadis (2009) used AOL query logs to analyze users' query reformulation strategies. Unlike the present study, they focused on clicks as a measure of success. They found that certain strategies resulted in more clicks – adding words, removing words, substituting words, expanding acronyms, and correcting spelling. However, as they addressed in their discussion, using query logs provides great volume, but misses the richness of interviews, where questions about users' satisfaction with results (and their reasoning behind the clicks) could be explored.

In their 2010 paper, Liu, Gwizdka, Liu, Xu, and Belkin explore the influence of task type and task situation on users' query reformulation behavior. They performed a controlled experiment with 48 participants. This controlled design is similar to the work closely related to this master's paper. They categorized tasks into three structures: simple, hierarchical, and parallel. This categorization is different that the one developed by Kelly et al. (2015) and eventually used in this paper, but it still speaks to a focus and interest in how different tasks have different effects on user behavior. The coding scheme used in Liu et al.'s study is similar in many ways to the scheme used in this work. They categorize reformulation into five types: generalization, specialization, word substitution, repetition, and new query. The scheme used in the present work is described in greater detail below. They found that specialization was a more common reformulation move in simple and hierarchically structured tasks, and that word substitution (i.e., using synonyms or trying new phrasing for the same concepts) was most common is parallel tasks.

1.3 Closely Related Work

This research described in this paper is a secondary analysis of data collected in Brennan et al. who measured participants' cognitive abilities (associative memory, perceptual speed, and visualization skill) using a series tests from the Manual for the Kit of Factor-referenced Cognitive Tests, and their impact of experienced workload during search. Because these data are used as the basis for this study, I will describe the method and findings of the study below.

For each of the three abilities, participants took a practice test and two actual tests, whose scores were added together to calculate the final score. Tests were timed. Participants were divided into low and high groups using a median-split for each of the three abilities. Associative memory was measured using the Picture-Number Test (MA-1). Perceptual Speed was measured Number Comparison Test (P-1). Visualization skill was measured using Paper Folding Test (VZ-2). The series of tests took approximately 28 minutes total. Workload was measured with the NASA Task Load Index (TLX), which assessed mental, physical, and temporal demands. The TLX looks at participants' perceived performance, perceived effort, and self-expressed frustration (Hart 1988).

Brennan et al. (2014) recruited twenty-one participants. There were 13 females and eight males, with an average age of 45.4. Nine identified as White, nine as Black, two as

Hispanic (ethnicity) and one Indian. All 21 held high school diplomas, with four holding associate's degrees, three holding bachelor's degrees, and three holding master's degrees. Their computer experience ranged from seven to over 10 years. All 21 had regular computer access, and reported using the Internet either daily (n=17) or 2-3 times per week (n=4).

After completing a demographic questionnaire and a series of cognitive ability tests, participants each completed their six assigned search tasks, given one at a time, and used copy and paste or manual entry to record their responses in a Microsoft Word document. Search behavior data, including queries, session length, SERP clicks, and dwell time, were recorded via transaction log during the session.

Using mixed model ANOVAs, looking at cognitive ability, task complexity, and perceived workload, Brennan et al. (2014) found that task complexity had a significant effect on search behavior. The more complex the task, the longer the search session, the more queries entered, the longer the queries, and the more SERP clicks per session. More complex tasks were also associated with lower performance, more mental demand, more physical demand, more temporal demand, high levels of frustration, and more perceived effort.

There were no main effects observed for associative memory. For visualization, there were no main effects on workload, but those in the "high" group entered significantly more queries than the low group, abandoned more queries, performed more SERP clicks, and looked at more URLs than the low group. Of the three cognitive abilities measures, perceptual ability had "the greatest effect on search behavior and workload" with the "low group experiencing greater demands than those in the high group" and engaging in more interaction while searching.

Methods

This paper represents a secondary analysis of data collected in Brennan et al.'s (2014) study. Each of the queries recorded during Brennan et al.'s study were coded using a scheme developed through personal communications over the course of spring semester 2015 (B.Wildemuth, D. Kelly, E. Boetcher, G. Abernathy, March 2015). The queries were broken down into individual search moves. Almost all of the queries included multiple search moves. For example, if a user searched for "skydiving risks" that would represent two "add concept" moves ('skydiving', as a representation of the concept 'Extreme Sports', and 'risks' as a representation of the 'Risk' concept). Appendix 1 provides the full text of the search tasks, along with the list of concepts for each. Table 1 provides an in depth description of the coding scheme used for this analysis

Due to time constraints, I did not perform any inter-coder reliability testing. Any future work in this area should attempt to develop a larger data set, and to use multiple coders to increase the reliability of the coding. In that situation, Cohen's Kappa could be calculated to judge the extent to which coders are in agreement, or else consensus-based coding would be used.

For the present paper, Brennan et al.'s high and low group assignments for each of the three cognitive abilities were recalculated, using the method described in the original work (Brennan et al., 2014). Of the 21 participants, one did not have associative memory test scores, and one did not have visualization test scores. For this reason, the sample size for the perceptual group is 21 while the visualization and associative memory have a sample size of 20. This resulted in each cognitive group having a slightly different total number of search moves. Nine out of 20 participants were in the high associate memory group, 10 out of 21 were in the high perception group, and 10 out of 20 were in the high visualization group.

The query reformulation codes were then analyzed based on overall frequency of occurrence as well as frequency based on cognitive ability levels (associative memory, perceptual speed and visualization skill). The results were evaluated using chi-square tests for independence and independent-samples t-tests with alpha = .05).

Table 1. Coding Scheme: Codes and Definitions

Exhaust	Include all the facets of the task; used only on the first query for a task
Replace term	Replace a term with a sibling/cousin term (i.e. a synonym or closely-related term) for the same concept; does not necessarily require a 1:1 equivalence
Add concept	Add a concept that is not represented in the previous search cycle (including the first query)
Narrow term	Replace a term with a narrower term for the same concept
Add Term	Add a term leading to a narrower specification of the concept
Delete Term	Remove a term without replacing it with a lateral, broader or narrower term
Broaden Term	Replace a term with a broader term for the same concept
Delete Concept	Moving to a broader concept by deleting terms or adding broader terms
Edit	Make minor (i.e., non-conceptual) changes in the query; includes moves originally coded as Rearrange (changing term order), Respace (spacing variants), Correct (correction of spelling errors), Add stop word, and Remove stop word
Error	Query that should be skipped when considering consecutive moves; applies only when the entire query should be considered an error. (Code the query as an error, then code the next query in relation to the query before the error)

Coding Scheme developed during series of personal communications (B. Wildemuth, D. Kelly, G. Abernathy, E. Boettcher, March 2015). Adapted from Wildemuth (2004), Shute and Smith (1993).

Results

There were 629 total moves made during the search tasks (Table 2). The most common move was *add concept*, representing 58.7% of moves. While the original coding scheme distinguished between adding terms, deleting terms, and replacing terms, for analysis purposes these codes were collapsed into *term manipulation*, which represented 18.8% of moves, followed by *delete concept*, which accounted for 14.6% of moves. *Edit, Error,* and *Exhaust* are included in Table 1 for reference, though they have been excluded from later analysis due to the infrequency of their occurrence. Table 2 shows the frequency of moves (all participants regardless of cognitive group).

Table 2. Frequency distribution of search moves (all participants).

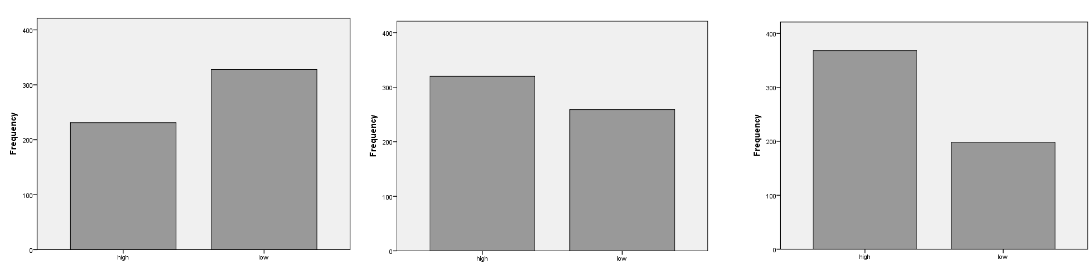
	Frequency	Percent	Cumulative Percent
add concept	369	58.7	58.7
delete concept	92	14.6	73.3
edit	14	2.2	75.5
error	14	2.2	77.7
exhaust	22	3.5	81.2
term manipulation	118	18.8	100.0
Total	**629**	**100.0**	

Figure 1 breaks these frequencies down according to the participant's associative memory score. Out of 559 total moves, the high associative memory group made 231, while those in the low associative memory group made the remaining 328 moves. An independent-samples t-test was conducted to compare the total number of moves for participants in the high and low associative memory groups. There was no significant difference in the scores for low (M=31.6, SD=19.3) and high (M=28.6, SD=11.9) groups; *t*(18)=.406, *p*=.69.

Figure 2 shows move frequency broken down by perceptual speed group. Out of 579 total moves, those in the high perceptual speed group made 320 moves, while the low group was responsible for the remaining moves. An independent-samples t-test was conducted to compare the total number of moves for participants in the high and low perceptual speed groups. There was no significant difference in the scores for low (M=25.5, SD=16.9) and high (M=34.9, SD=13.3) groups; *t*(19)=-1.415, *p*=.173.

In Figure 3, the same analysis takes place, but broken down according to visualization ability. Out of 566 moves, 368 were those in the high group, the other 198 by those in the low group. Visualization skill has the most drastic difference between the two groups. An independent-samples t-test was conducted to compare the total number of moves for participants in high and low visualization groups. There was a significant difference in the scores for low (M=21.8, SD=5.181) and high (M=39.7, SD=17.62) groups; *t*(10.5)=-3.082*, p*=.011. In this experiment, those in the high visualization group as a whole performed significantly more query reformulation moves than those in the low visualization group.

Figures 1, 2, 3. Total number of search moves for high and low groups. Left to right: associative



memory, perceptual speed, and visualization skill.

Finally, three chi-squared tests of independence were performed, to look at the relationship between cognitive ability group and the distribution of search moves by type. The number of search moves by type as a function of associative memory group is shown in Figure 4. The difference in frequencies was not significant, $X^2(2, N=559) = 2.893$, $p=.235$. The number of search moves by type as a function of perceptual speed group is shown in Figure 5. The difference in frequencies was not significant, $X^2(2, N=579) = 3.648$, $p=.161$. The number of search moves by type as a function of visualization group is shown in Figure 6. The difference in frequencies was significant, $X^2(2, N=566) = 27.979$, $p<.001$. The largest difference is in the number of term manipulation search moves. Those in the high visualization group made more term manipulation reformulations than those in the low group.

Figure 4. Frequency of search moves by type and associative memory group.

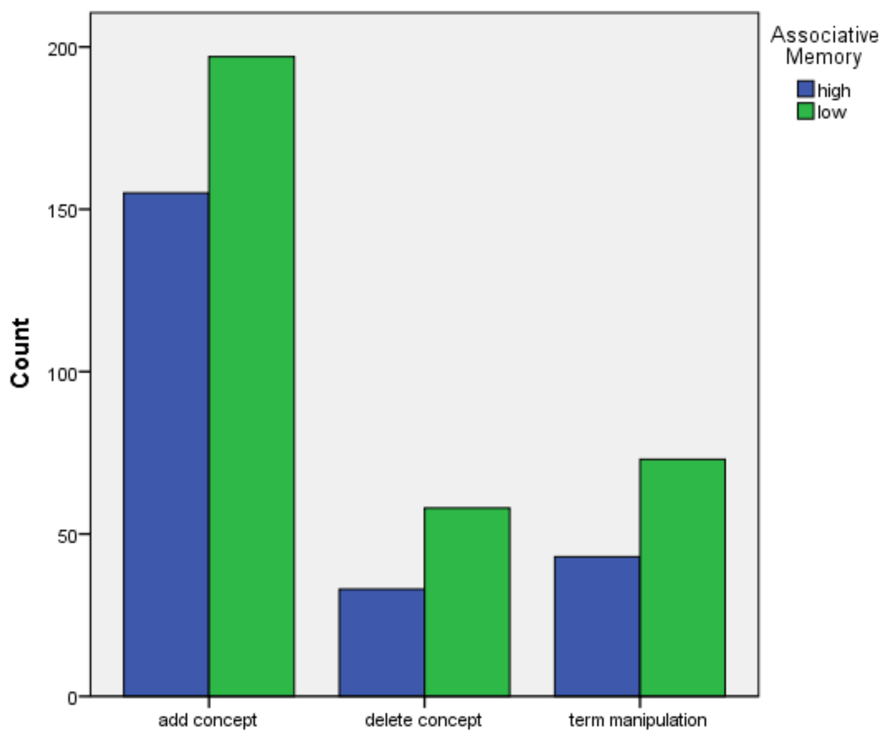


Figure 5. Frequency of search moves by type and perceptual speed group

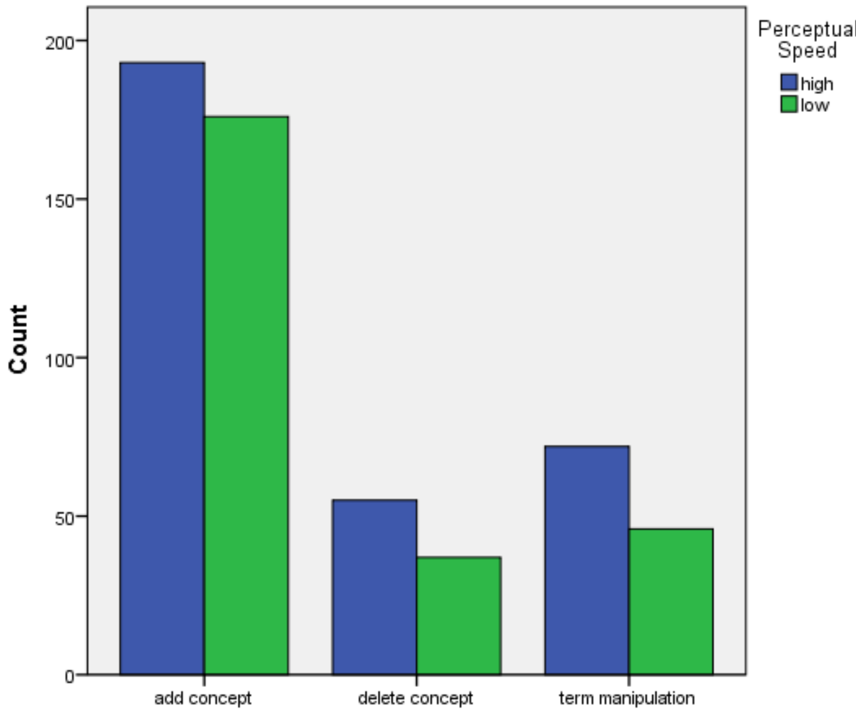
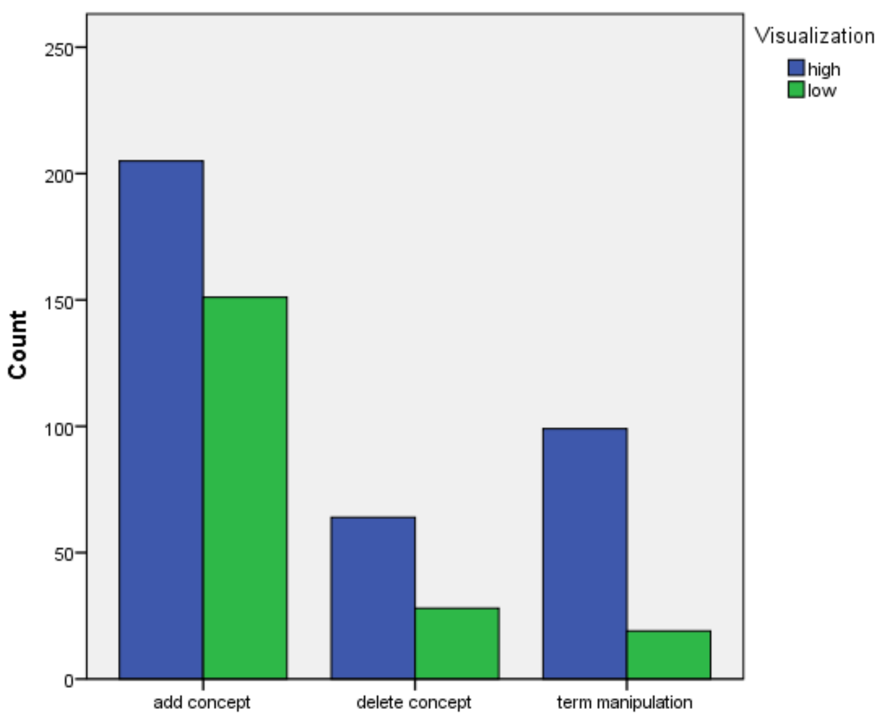


Figure 6. Frequency of search moves by type and visualization skill group.



Discussion

The significant result of this secondary analysis is that those in the high visualization ability group made significantly more total search moves, and significantly more term manipulation moves, than did those in the low group. What could this indicate? Brennan et al. (2014) note that visualization

> requires the ability to situate one's self in relationship to a static object, such as being able to imagine a piece of paper in its various stages from being folded to the end, being completely unfolded […] The ability to think sequentially, with a strategy. (p. 167)

Downing et al. (2005) found that those with low visualization abilities found fewer relevant documents, and took longer to find those documents. Perhaps the total number of moves observed in this work indicates that high visualization means more ideas, or a clear search plan, with sequential awareness of what they've already tried, and what they want to try next. Similarly, the increase in term manipulation moves might suggest that those with higher visualization ability can conceptualize the various elements of a search task, and substitute synonyms, narrower terms, or broader terms, without losing sight of the various elements of the task at hand.

Years earlier, however, Swan and Allan (1998) did not find a relationship between visualization skill and ability at interacting with a 3-D interface. The novelty of the interface may have impacted the results. Given that past research does not present a consensus on the role of visualization in search success and search behaviors, the results of this paper add to the body of work on this topic. Previous work judged search success on number of found results, time spent searching, or time to first relevant results. This paper looks at different measures, and so is not directly comparable to past work on this topic.

Despite its findings, this paper has several limitations. The first is a lack of inter-coder reliability testing, and a lack of multiple coders. Due to time constraints, I coded all of the queries myself. While experienced using the same coding scheme on different query data, which was discussed in great depth with other coders, that is insufficient. Any elaboration on the work outlined in this paper should have multiple coders, either working towards consensus, or being tested against each other.

Another factor that separates this work from its predecessor is the use of display data. That is, information about what users clicked on, not only from the SERP page(s), but any additions browsing clicks. While this information does not invalidate the analysis, its addition would allow a more complex look at search formulation behaviors and patterns, and might help establish relationships between moves made and sites clicked (and dwell time on those sites).

Some anecdotal observations during coding and analysis present interesting opportunities for future research. A substantial minority of recorded queries were written using natural language: using articles, adverbs, conjunctions, etc. rather than just keywords or Boolean syntax. While the coding scheme used in this paper did not distinguish differences in syntax, it would be interesting to look at sentence-like queries in conjunction with both cognitive ability and workload. Are participants more likely to use natural language when they perceive the task as complex, difficult, or unclear?

Similarly, some participants utilized Boolean syntax, despite Google and other major search engines neither requiring nor encouraging it.

Another behavior that occurred infrequently but was of interest was the use of "concept 0." In coding queries, numbers were assigned to the various facets of a task. If the participant used a keyword that did not fall into any of the facets, it was coded as "concept 0." It would be interesting, particularly on a larger scale, to see if this behavior correlated with cognitive ability, task complexity, or any of the other elements measured in related papers. On the one hand, you might hypothesize that "concept 0" indicates a creative mind with a strong grasp of the task; on the other, it might indicate confusion, frustration, or lack of search formulation ability.

In a similar vein, many participants pulled words and phrases from the assigned search tasks verbatim. Often, they used parts of the task description that were part of the context but not part of the question. While my initial sense is that they simply started searching before reading to the end of the question, without coding for this behavior, all I can report is a hunch. It would be interesting to look at cognitive ability in this context, as reading comprehension and the ability to pull out keywords are popular test taking skills. There would likely be educational applications for such research.

One participant's searches stood out to me. The experiment allowed the participant to use major search engine. Most used Google, but one navigated to Bing at the start of each query. While a single user is anecdotal in the scope of this paper, it would be interesting, once relationships have been observed regarding cognitive ability, perceived workload, task complexity, and user engagement, to then add search engine preference as one more variable for analysis.

Conclusion

Three facets of cognitive ability were analyzed in this study. I explored how associative memory, perceptual speed, and visualization abilities related to total number of search moves, as well as the distribution of those moves by type.

Preliminary findings suggest a relationship between visualization skill and number of moves, particularly term manipulation moves. Those participants who made up the high visualization skill group made significantly more formulation and reformulation than their counterparts.

This paper's goal was to build on previous work, combining dimensions of search strategy formulation and reformulation with assessment of cognitive abilities. Reformulation strategies, when taken alongside assessment of task complexity, perceived workload, and task domain, may aid in the development of search tasks for further IIR research.

Additional research might take larger samples, or incorporate qualitative interviews, to explore the impact of visualization skill on search behavior and search skill. Work in that vein could lead to better understanding of information search needs, as well as to the development of more effective search engine interfaces for a variety of everyday users.

Acknowledgements

I would like to express my gratitude to professors Diane Kelly, Barbara Wildemuth, and Jaime Arguello. In addition, thanks to Kathy Brennan, Ashlee Edwards, Wan-Ching Wu, Emma Boetcher, and Gergana Abernathy. Without their previous work, this paper would not have been possible.

References

Al-Maskari, A., & Sanderson, M. (2011). The effect of user characteristics on search effectiveness in information retrieval. *Information Processing & Management*, *47*(5), 719-729.

Allen, B. (1994). Perceptual speed, learning, and information retrieval performance. In *Proceedings of SIGIR conference on Research and development in information retrieval,* 71-80. ACM.

Brennan, K., Kelly, D., & Arguello, J. (2014, August). The effect of cognitive abilities on information search for tasks of varying levels of complexity. In *Proceedings of the 5th Information Interaction in Context Symposium* (pp. 165-174). ACM.

Campagnoni, F.R., and Erlich, K. (1989). Information retrieval using a hypertext-based help system. In *Proceedings of ACM SIGIR conference on Research and development in information retrieval* (pp. 212-220).

Downing, R.E., Moore, J.L., and Brown, S.W. (2005). The effects and interaction of spatial visualization and domain expertise on information seeking. In *Computers in Human Behavior, 21(2),* 195-209.

Ekstrom, R.B., French, J.W., Harman, H.H., and Dermen, D. (1976). *Manual for Kit of Factor-Referenced Cognitive Tests.* Educational Testing Service, Princeton, NJ.

Gwizdka, J. (2009). What a difference a tag cloud makes: effects of tasks and cognitive abilities on search results interface use. In *Information Research, 14(4),* paper 14.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*, *52*, 139-183.

Huang, J., & Efthimiadis, E. N. (2009, November). Analyzing and evaluating query reformulation strategies in web search logs. In *Proceedings of the 18th ACM conference on Information and knowledge management* (pp. 77-86). ACM.

Kelly, D., Arguello, J., Wu, W.C. and Edwards, A. (2015). Development and evaluation of search tasks for IIR experiments using a cognitive complexity framework. *Proceedings of the ACM International Conference on the Theory of Information Retrieval (ICTIR '15),* Northhampton, MA

Liu, C., Gwizdka, J., Liu, J., Xu, T., & Belkin, N. J. (2010). Analysis and evaluation of query reformulations in different task types. *Proceedings of the American Society for Information Science and Technology*, *47*(1), 1-9.

Pak, R., Rogers, W.A., and Fisk, A.D. (2006). Spatial Ability Subfactors and Their Influence on a Computer-Based Information Search Task. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 48(1),* 154-165.

Rieh, S. Y., and Xie, Hong (Iris). (2006). Analysis of multiple query reformulations on the web: The interactive information retrieval context. *Information Processing & Management*, *42*(3), 751-768.

Shute, S. J. and Smith, P. J. (1993). Knowledge-based search tactics. *Information Processing & Management*, *29*(1), 29-45.

Spink, A., Jansen, B. J., & Cenk Ozmultu, H. (2000). Use of query reformulation and relevance feedback by Excite users. *Internet research*, *10*(4), 317-328.

Swan, R. C., & Allan, J. (1998, August). Aspect windows, 3-D visualizations, and indirect comparisons of information retrieval systems. In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 173-181). ACM.

Westerman, S.J., Davies, D.R., Glendon, A.I., Stammers, R.B., and Matthews, G. (1995). Age and cognitive ability as predictors of computerized information retrieval. In *Behavioral Information Technology, 15(5*), 313-326.

Wildemuth, B. M. (2004). The effects of domain knowledge on search tactic formulation. *Journal of the American Society for Information Science and Technology*, *55*(3), 246-258.

Appendix 1: Search Task Descriptions and Concepts

Tasks by Kelly et al. (2015), list of concepts developed through personal communications (B. Wildemuth, D. Kelly, E. Boetcher, G. Abernathy, March 2015).

Science & Technology/Remember
You recently watched a show on the Discovery Channel, about fish that can live so deep in the ocean that they're in darkness most or all of the time. This made you more curious about the deepest point in the ocean. What is the name of the deepest point in the ocean?
Concepts:

1. Deepest point
2. Ocean

Science & Technology/Analyze
You recently became involved with a conservation group that picks-up trash from local waterways. One of the group members told you that your work was important because it helps keep pollution out of the ocean. What are some of the different types of ocean pollutants? What environmental risks are associated with each pollutant?
Concepts:

1. Pollutant(s)
2. Ocean
3. Types of / Varieties
4. Environmental
5. Risk

Science & Technology/Create
After the NASCAR season opened this year, your niece became really interested in soapbox derby racing. Since her parents are both really busy, you've agreed to help her build a car so that she can enter a local race. The first step is to figure out how to build a car. Identify some basic designs that you might use and create a basic plan for constructing the car.
Concepts:

1. Derby (race)
2. Car (soapbox)
3. Design
4. Construction/building

Entertainment/Remember
You recently attended an outdoor music festival and heard a band called Wolf Parade. You really enjoyed the band and want to purchase their latest album. What is the name of their latest (full-length) album.
Concepts:
1. Wolf Parade
2. Album Name (includes "album")
3. Album Length (EP vs LP vs Demo)
4. Latest / Newest / Date term

Entertainment/Analyze
Your sister is turning 25 next month and wants to do something exciting for her birthday. She is considering some type of extreme sport. What are some different types of extreme sports in which amateurs can participate? What are the risks involved with each sport?
Concepts:
1. Extreme sports (includes specific kinds of extreme sports)
2. Amateur
3. Risks
4. Types, lists, etc.

Entertainment/Create
Your local Triple-A affiliate baseball team has decided that it is time for a new mascot and are holding a contest where fans can enter suggestions. Being a loyal fan, you have decided to enter the contest. You want to suggest a mascot that will appeal to many people and will represent important qualities of a baseball team. The team is a part of the International League, so you want to avoid suggesting a mascot that is already represented in this league. Which mascot would you pick and why?
Concepts:
1. Mascot
2. Baseball
3. League (including specific league and that it's a Triple-A league)
4. Non-duplication
5. Appeal to fans
6. Representative of (important) team qualities
7. Localness